VISUAL QUESTION ANSWERING IN THE MEDICAL DOMAIN (VQA-MED)

Volviane Saphir MFOGO;Xinlei Chen;Jeremiah Fadugba;Georgia Gkioxari;

;

IDRC Grant/ Subvention du CRDI: 109187-002-Laying the foundations for artificial intelligence for development (AI4D) in Africa

Visual Question Answering in the Medical Domain (VQA-Med)

Volviane Saphir MFOGO

Xinlei Chen

Jeremiah Fadugba

Georgia Gkioxari

Contents

- Motivation of VQA-Med and Benefit
- VQA-Med dataset
- Methodology
- Result and Discussion
- Next steps

Motivation of VQA-Med

- The need for medical AI models

Cardiologist

Gynecologist

Radiologist

Motivation of VQA-Med

- Disappointed during hospital visit

Benefit of VQA-Med in Africa

In this project we aim to build a medical AI assistant which has the potential to complement clinician's diagnoses. It will be benefit to African community, Hospitals and Researchers

African community and hospitals

Used as “second opinion” to Medical doctor and patient as well.

Researchers

- Researchers in Africa and beyond could use it as a baseline.
- This could also encourage and open the path for more specialized data collection and the start of more in-depth research in the field of health.

DEEP LEARNING INDABA

VQA-Med Dataset Description

- Applying several filters to select relevant images and associated annotations;
- Creating patterns to generate the questions and their answers.
- The test set was manually validated by two medical doctors.
- The dataset is publicly available.

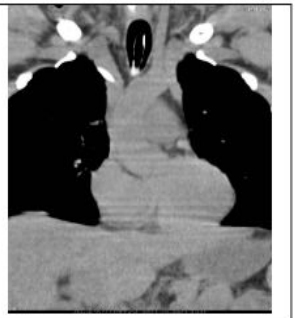
Image		
Question	what modality was used to take this image?	in what plane was this image taken?
Answer	mammograph	coronal

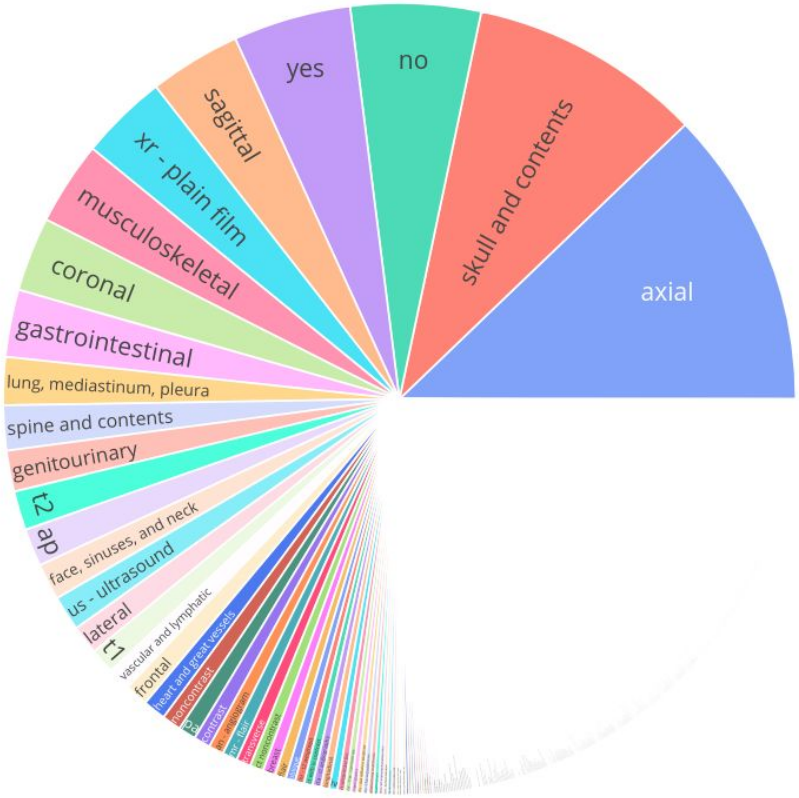
Image		
Question	which organ is captured by this mri?	what is the primary abnormality in this image?
Answer	skull and contents	autosomal recessive polycystic kidney disease

Dataset examples for each category

VQA-Med Dataset Anomalies

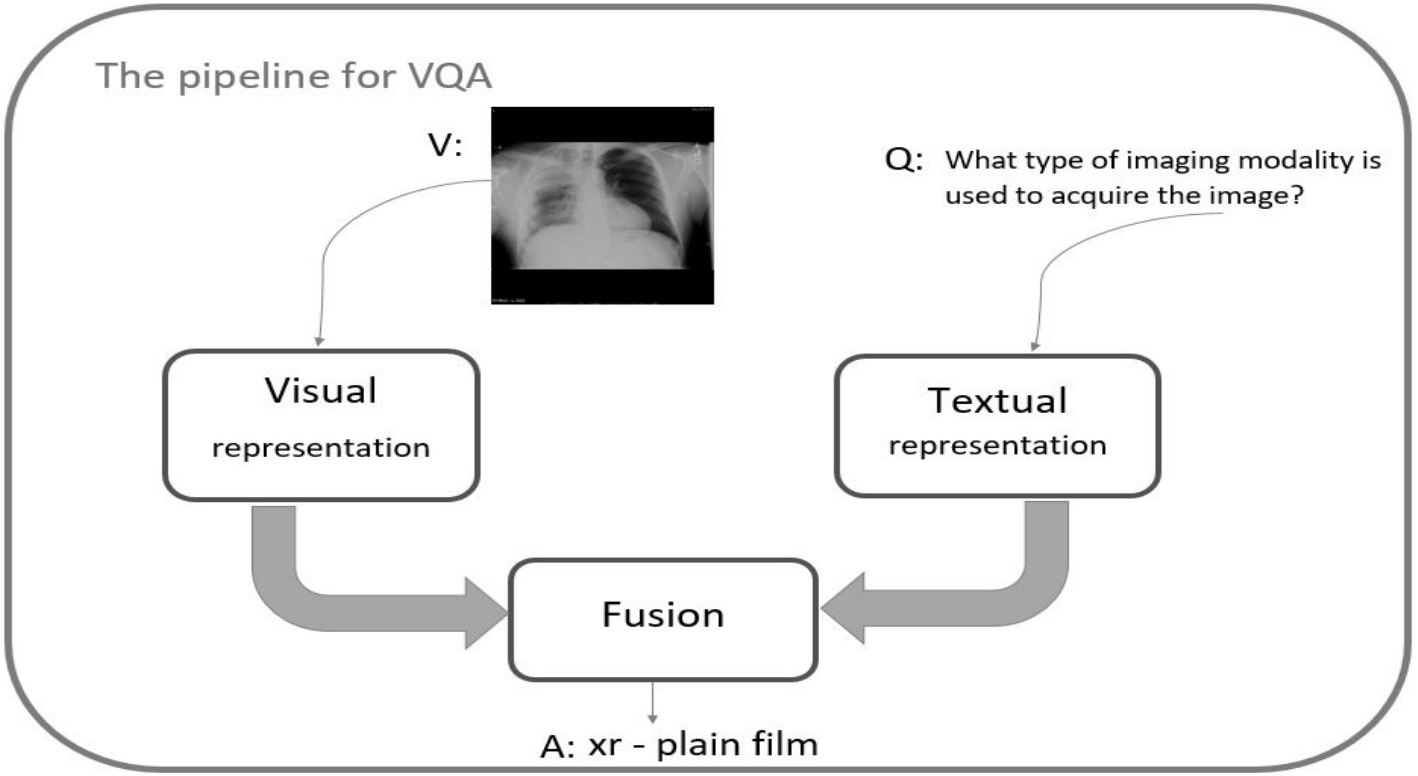
The dataset is from [VQA-Med-2019] which includes:

- The training set of 12792 question-answer pairs associated with 3200 medical images.
- The validation set of 2000 question-answer pairs associated with 500 medical images.
- The test set of 500 question-answer pairs associated with 500 medical images.



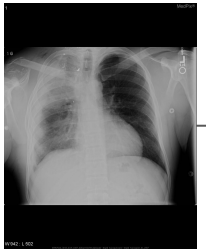
- The dataset is too small for exploring advanced VQA capabilities;
- Imbalance in the answer;

Methodology



The pipeline for VQA

V:



Q: what type of imaging modality is used to acquire the image?

VGG +GAP

BERT

Image feature att.

MFB

Question feature att.

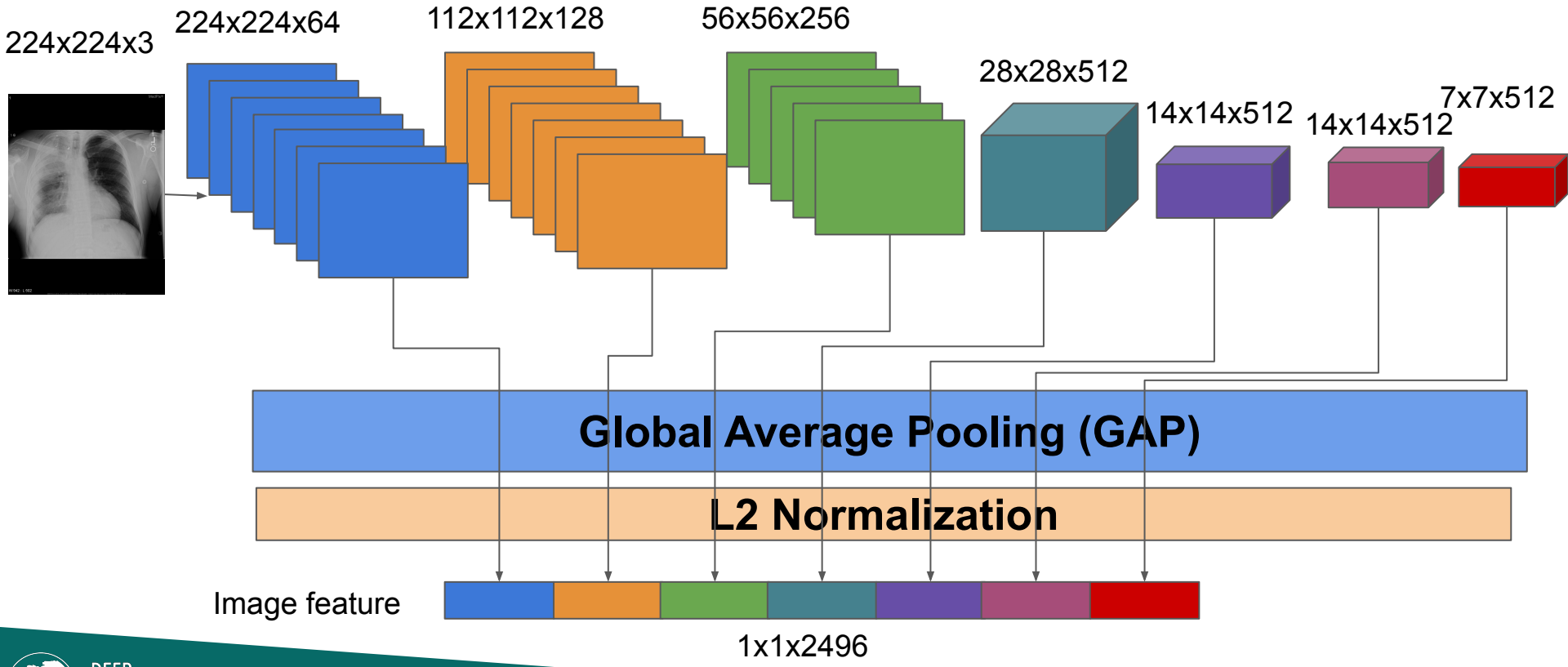
MFB

fc

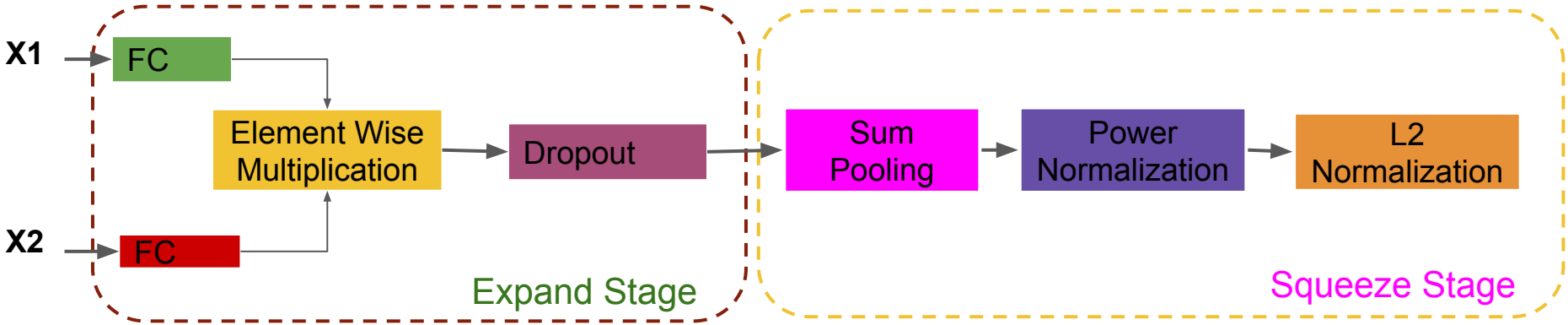
Softmax

A: xr - plain film

- Backbone : VGG19



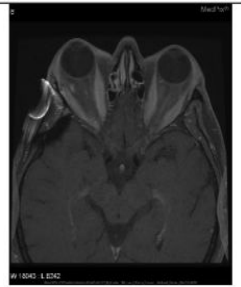
- Multi-modal Factorized Bilinear pooling (MFB)



Result & Discussion

Giving Image on test		
Giving Question on test	what modality is shown?	in what plane is this ct scan?
Model Answer prediction	cta - ct angiography	axial

Model prediction

Giving Image on test		
Giving Question on test	what organ system is displayed in this ct scan?	is there an abnormality in the mri?
Model Answer prediction	skull and contents	yes

DEEP LEARNING INDABA

Result & Discussion

Category	Modality	Plane	Organ system	Abnormality	Overall
Accuracy	0.858	0.710	0.669	0.081	0.579

Model accuracy

- To deal with that imbalance, We consider only the classes that occur a least 5 times, the rest we group them in the same class call ‘UNKNOWN’.

- The bad performance of our model on Abnormality question is due to the fact that the number of class of this category is very large (1485 classes), where we have a lot of class that appear on the training dataset only once.

Next Steps

- Improve the visual and/or fusion side of the model
- Explore more model architecture
- Make the code reproducible and open source it
- Write the technical paper